DOCUMENT RESUME

ED 318 748 TM 014 761

AUTHOR Waters, Brian K.
TITLE Development of a Single DoD Reading Grade Level (RGL) Scale.
PUB DATE Aug 89
NOTE 30p.; Paper presented at the Annual Meeting of the American Psychological Association (97th, New Orleans, LA, August 11-15, 1989). Information was drawn from "Estimating the Reading Skills of Military Applicants: Development of an ASVAB to RGL Conversion Table" by B. K. Waters et al. (1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Equated Scores; *Military Personnel; Military Training; Rating Scales; Reading Achievement; Reading Skills; *Reading Tests; Statistical Analysis; *Test Construction
IDENTIFIERS Anchor Tests; Armed Services Vocational Aptitude Battery; Department of Defense; *Reading Grade Level Scale

ABSTRACT

Research on military reading abilities has shown a direct relationship between reading ability and training and job performance. The objectives of this study of the development of a reading scale were to: (1) measure the reading ability of military applicants using six commonly used reading tests; (2) use the Armed Services Vocational Aptitude Battery (ASVAB) as a composite anchor against which the reading tests can be equated; (3) equate each test to the anchor; and (4) recommend a reading grade level (RGL) scale for Department of Defense (DoD) use in reporting to the U.S. Congress. Subjects were 20,422 applicants for the armed forces who were tested during a 6-week period beginning on April 13, 1987 on the ASVAB and one of the 6 reading tests. Scores on the reading tests yielded widely varying average RGL levels for examinees of equivalent ability. The ASVAB verbal standard score was taken as the anchor, and a conversion table was made for the ASVAB and each of the five reading tests found adequate for equating purposes. An RGL scale was then proposed for DoD use that would allow each branch of military service to use the reading test that best meets its requirements, with each converting the scores to the RGL scale for congressional reporting purposes. Thirteen tables and eight graphs present the study results. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED318748

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BRIAN K. WATERS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DEVELOPMENT OF A SINGLE DoD READING GRADE LEVEL (RGL) SCALE

Brian K. Waters
Human Resources Research Organization (HumRRO)

TM014761

BEST COPY AVAILABLE

DEVELOPMENT OF A SINGLE DoD READING GRADE LEVEL (RGL) SCALE [1]

Brian K. Waters
Human Resources Research Organization (HumRRO) [2]

Research has consistently shown that there is a strong relationship between reading ability and job performance in the military. Thus, literacy levels are of great concern to those responsible for setting selection, classification, and training policy. The need for greater reading skills has become increasingly more evident as military jobs have become more complex.

The level of basic skills of entering personnel has been a key military manpower consideration for some time. In fact, before the advent of the Armed Forces Qualification Test (AFQT) selection was often based primarily upon literacy. During World War II, for example, the military used "ability to read and write English at the fourth grade level" (Eitelberg, Laurence, & Waters, with Perelman, 1984, p. 24) as the minimum level of literacy. As military jobs have become more technologically complex they require higher reading and computational skills. Today, working for your country demands considerable basic skills levels and there is strong interest in how many recruits "stack up." This interest extends beyond the Pentagon mazes. Congress is concerned with the "quality" of recruits and their representativeness relative to the national population of youth. To address these concerns, Congress requested DoD to report recruit characteristics such as high school graduation status, enlistment aptitude test scores, RGL, race, sex, geographic region, and age. Of particular importance is the reading ability of recruits. (DoD, 1988)

RGL Scale Development

RGL scales ideally are developed through a large-scale standardized testing within all grades from first through post-high school. The samples should be nationally representative of all students at each grade level. Particularly at post-secondary levels, this requirement is very difficult to meet, resulting in atypical samples such as vocational/technical students or prison populations used as the basis for norming. In many cases, reading

---

[1]Paper presented at the 97th Annual Convention of the American Psychological Association, Division of Military Psychology, August, 1989, New Orleans, Louisiana. Opinions expressed are those of the author, and do not reflect Government policy or position unless published in official documentation.

[2]Information in this paper was drawn from a HumRRO Technical Report, Waters, B. K., Barnes, J. D., Foley, P., Steinhaus, S. D., & Brown, D. C. (1988). Estimating the Reading Skills of Military Applicants: Development of an ASVAB to RGL Conversion Table (HumRRO FR-PRD-88-22). Alexandria, VA: Human Resources Research Organization.

TM 014761

tests establish their RGL scales through "equating" of scores via a sample of examinees who take two reading tests: their own and a previously normed "anchor" test.

Different reading tests have their own norming groups, thus a person's RGL on one test may be quite different from the score of the same person on a different test of reading ability. Differences in scale development are not the only factors which make it difficult to compare RGL scores from different tests. The concept of RGL is often not well understood and the scores are misused. In theory, an RGL score indicates the grade and month in school for which a given test score is average. For example, if the mean raw score of students in the third month (November) of the ninth grade in the norming sample for a given reading test is 27, then all persons scoring a 27 would be assigned a RGL of 9.3. However, school testing normally occurs twice during the school year -- in the Fall (generally October) and in the Spring (generally April). At each grade level, mean scores are calculated and then a smoothed extrapolation and interpolation of scores is made. For example, if the mean of Fall testing is 27 (9.3) and Spring is 34 (9.9), then the RGL in January would be 9.6. For some tests, such as the Nelson-Denny Reading Test, the scale runs through a RGL of 16.9; although the meaning of a post-college graduate-level reading ability is not clear.

In addition, testing and statistical "experts" point out that the underlying scale is not interval, nor really even monotonic; that is, the reading ability of an average fifth grader in May (5.9) is consistently higher that the average sixth grader the following September (6.0). Teachers suggest that a loss of tested reading ability occurs over the summer break, and that the early part of the school year brings the students back to their level of proficiency in the previous Spring. Such vagaries have led RGL critics to recommend that the RGL scale be eliminated; however, it seems highly unlikely that this will occur. Thus, it becomes essential that test developers appropriately counsel test users on proper interpretation of the RGL scale. For example, it should be pointed out that reading scores vary rather widely around a school year's mean RGL. Individual differences exist in both reading ability and in learning rates. A typical third grade class might have student RGL estimates from first through fifth grade levels for individual students, and although a 1.1 to 2.1 RGL change would be "normal" on the average between October of the first and second grades, brighter students might be expected to increase more and less bright students less than one RGL over a one-year period.

Military Reading Research

Research on military reading abilities shows a direct relationship between reading ability and training and job performance. Thus, it appears that developing a new RGL scale for the Armed Services Vocational Aptitude Battery (ASVAB) that would be both current and relevant to military uses--selection and classification, development of training materials, and conducting remedial training programs -- would allow the Services to man their forces more efficiently. It would also facilitate reporting of recruit quality to the Congress.

Study Objectives

The objectives of this study were to: 1) measure the reading ability of military applicants using six commonly used reading tests; 2) select one ASVAB composite (anchor) against which to equate the reading tests; 3) equate each reading test to the anchor; and 4) recommend a RGL scale for use by DoD to report recruit reading ability to the Congress. The following sections of this paper present the approach, results, conclusions, and recommendations of this study.

APPROACH

Subjects

The subjects were 20,422 applicants for the Army, Navy, Air Force, Marine Corps, and Coast Guard (including applicants for the Reserves and National Guard). They were tested on the ASVAB and one of six reading tests during a six-week period beginning April 13, 1987. Each of the reading tests was administered to more than 3,000 applicants.

Testing took place at 15 of 68 Military Entrance Processing Stations (MEPSs) and at 214 of the nearly 900 Mobile Examining Team (MET) sites which are subordinate to the MEPSs. Both MEPSs and MET sites fall under the U. S. Military Entrance Processing Command (USMEPCOM). Applicants at the MEPS receive both ASVAB testing and medical testing and are processed for enlistment. The MET sites, on the other hand, provide only ASVAB testing. They conduct over 80 percent of ASVAB testing. Successful MET site examinees proceed to the MEPSs, where the balance of their enlistment screening occurs.

Instruments

ASVAB Information

The RGL study used then operational ASVAB Forms 11, 12, and 13. The ASVAB is a multiple aptitude test battery measuring general learning ability and potential in occupations which require formal courses of instruction or on-the-job training.

Research has documented the ASVAB as a reliable and valid multi-factor aptitude test battery. Reliabilities of the ASVAB composites evaluated in this study were high -- .87 to .95. Service personnel research laboratories have conducted an impressive collection of studies of the development and validation of the ASVAB.

Reading Test Information

Table 1 contains the names of the six reading tests used in this study, the level and form used, and the populations for which they were designed.

Table 1

Description of RGL Study Reading Tests

Reading Test	Level	Form	For Grades
Adult Basic Learning Examination (ABLE)	3	E	9-12
Air Force Reading Abilities Test (AFRAT)	--	A	AF enlistees
Gates-MacGinitie (G-M)	F	1	10-12
Nelson-Denny (N-D)	--	E	9-12
Test of Adult Basic Education (TABE)	A	5	8-12
Stanford Tests of Academic Skills (TASK)	1	E	8-12

Each reading test contains separate vocabulary and comprehension subtests. All of the reading tests, except for the ABLE, also provide a total reading score. The reading tests had target populations of similar ages, but quite different norming groups cited in their test manuals.

ABLE: The Adult Basic Learning Examination (ABLE) Level 3, Form E was designed for grades 9 through 12 and published by the Psychological Corporation. The ABLE measures educational achievement in reading, mathemetics, and language arts. The ABLE does not report a total reading score, thus no total RGL norms were available. ABLE test items are suitable for adults, and present situations that are common to the everyday lives of adults. (ABLE Norms Booklet, Level 3, Second Edition, 1986, p.47). The Vocabulary test requires selection of one of three alternative words to complete a sentence. The Reading Comprehension item format uses relatively short prose passages, followed by four-choice questions on the content of the passage.

The ABLE consists of two sections. The first, vocabulary, contains 50 items which ask about the meaning of words in phrases. The second, comprehension, contains 58 items which involve questions about single paragraphs. The test requires approximately 50 minutes. Vocabulary, reading, and problem solving sections were used to calibrate the ASVAB general composite to ABLE in an Army study (Grafton, 1980). ABLE reports RGLs from 3.0 to 12.9.

Norming Groups: Test developers collected data during the winter of 1984, using three norming groups: 1) 515 adult prisoners; 2) 474 Adult Basic Education/General Educational Development high school equivalency program participants; and 3) 718 Vocational/Technical school students. The ABLE combines the three groups into a single norming group of 1,707 respondents.

The ABLE norming sample was derived from state mailing lists for adult education and vocational programs in communities and schools, and from mail-

ing lists for national, state, and local penal institutions. Every tenth program was requested to participate. The level given was determined by the most recent grade completed by the adult (grades 1-4 for Level 1, grades 5-8 for Level 2, and grades 9-12 for Level 3). An experimental form (A, B, C, or D) was selected by random assignment of one form to each program. Roughly 4,000 adults in 132 programs in 41 states participated in the research program. The groups consist of 59.8% male and 40.1% female and ages ranged from 19-29 (76.2%) to 50 years old. The racial/ethnic composition was white (54.3%), black (21.0%), Asian, Native American, and Hispanic (9.2%). Geographic regions were representative within the U. S. (Northeast, 20.0%; Midwest, 37.7%; Southeast, 27.5%; and West 14.8%).

AFRAT: Form A of the Air Force Reading Abilities Test (AFRAT), designed for Air Force enlistees, is published by the Air Force Human Resources Laboratory. AFRAT design stressed espository prose comprehension passages, with comprehension items unlikely to be answered correctly based solely on prior subject knowledge. Vocabulary words which were not likely to be encountered in the work environment, maximum reliability, and test length of less than one hour (Mathews & Roach, 1983; Valentine, 1980) were also design objectives. The vocabulary test item format uses a single word stem followed by four alternative synonyms. AFRAT Reading Comprehension passages are relatively long with four-choice response alternatives.

The AFRAT consists of vocabulary items in a multiple choice synonym format and comprehension items consisting of one or more paragraphs followed by one or more multiple choice questions. The comprehension items require either paraphrasing or making inferences from the passages. AFRAT forms A and B each contain 45 vocabulary and 40 comprehension items with a test limit of 15 minutes for vocabulary and 35 minutes for comprehension. The test was targeted at the 8th RGL as measured by the Adult Basic Learning Examination (ABLE). Although AFRAT Forms A and B were to be peaked at a difficulty level parallel to that of the 8th RGL, the preferred norms would span from the 5th through the 12th RGL.

Norming Groups: The AFRAT used seven samples of examinees tested from May to July 1981. The AFRAT was designed to measure Air Force recruits (primarily 18-23 years old). Six of the seven norming samples were Air Force recruits. One was an Army sample of 1,049 recruits (Mathews & Roach, 1983, pp. 2-3).

A total of 12,938 examinees were administered two reading tests, either two parallel forms of the AFRAT, A and B, or one AFRAT form and the Gates-MacGinitie Reading Test. Analyses were calculated to ascertain the resemblance of the AFRAT forms, their correlation with other reading tests, AFRAT raw scores to RGL equivalents, and training grade validity of AFRAT item types.

The AFRAT norms are based on data collected in earlier studies from the ABLE Grade L, TABE RGL, and G-M RGL. Specifically, the norms were established by averaging the RGLs across the aforementioned commercial tests.

G-M: Level F, Form 1 of the Gates-MacGinitie Reading Test (G-M), designed for grades 10 through 12, is published by Riverside Publishing Company. The G-M Vocabulary test has single word stems and four-response alternatives. The G-M Comprehension test has average length passages, followed by four choices. Overall, the G-M content and format are similar to the AFRAT.

Items of the G-M consist of 50 vocabulary and 42 comprehension with a combined testing time of 40 minutes. The vocabulary items require the selection of synonyms of single words and the comprehension items consist of questions about single paragraphs. RGLs of 1.0 to 11.9 are reported (Gates-MacGinitie, 1972).

Norming Groups: The G-M conducted a standardization using students from 1st through 12th grades in October 1976 and May 1977. Level F (10th through 12th grade) norms were based on four school samples totaling 15,160 respondents.

The G-M norms were obtained from a stratified sampling design based primarily on the Fourth Count of the 1970 U.S. Census, which provided data on the basis of school district boundaries. The norms were calculated at both the beginning (October, 1976) and end (May, 1977) of the school year. Eighty-six school districts, including parochial schools, participated in the norming phase of the standardization. Selection of districts was directed toward producing a representative proportion of black and Hispanic students. Norming samples included approximately 5,000 students per grade. The total number of students involved in each standardization testing was approximately 65,000. The normative scores--NCE scores, percentile ranks, stanines, grade equivalents, and extended scale scores--can all be derived directly from raw scores.

N-D: The Nelson-Denny Reading Test (N-D), Form E, developed for grades 9 through 12, is also published by Riverside. It is unlike the other reading tests or ASVAB subtests in format, difficulty, and speededness. The N-D Vocabulary test has a sentence stem with a single highlighted word in context, followed by five alternative responses. Each item is one line long, so the N-D Vocabulary test has many (25) items per page. The 100 item Vocabulary test has a 15 minute time limit; it requires many more items per minute than the other reading tests. Hence, the N-D Vocabulary test is highly speeded. The N-D Comprehension test has relatively short passages, also followed by five choices. The N-D allows 20 minutes for 36 items -- the shortest time per item of the comprehension tests. The N-D also yields a "reading rate" for each examinee.

Norming Groups: The N-D national standardization used three samples from November 1979: a high school sample of nearly 14,000 students representing both public and private 9th through 12th graders; a two-year college sample (n = 6,000); and a four-year college sample (n = 5,000). Forms E and F (9th to 15th grade) norms used the total 25,000 students.

School districts, including parochial schools, were sampled to represent all socioeconomic levels, geographic regions, and district sizes. The high

school sample consisted of nearly 14,000 students from 69 public school districts and 11 parochial schools. The two-year college and four-year college/university samples were selected on the basis of two criteria--the geographic region and the size. A total of 41 two-year colleges and 32 four-year colleges and universities participated in the standardization. The samples consisted of over 6,000 and 5,000 students, respectively.

TABE: Level A, Form 5 of the Test of Adult Basic Education (TABE) was designed for grades 8 through 12. It is published by CTB/McGraw-Hill. The TABE Vocabulary test contains four different four-choice formats: a two or three word phrase with one word highlighted which asks for a synonym or antonym; two phrase stems for which each asks for a common synonym; a single word in which a prefix or suffix is highlighted and the respondent selects the meaning of the highlighted part of the word; and a paragraph which includes sentences with a word omitted from the stem. The Comprehension subtest has four-choice response alternatives, referenced to relatively short passages of text.

The TABE consists of a 30-item vocabulary section and a 40-item comative proportion of black and Hispanic students. Norming samples included approximately 5,000 students per grade. The total number of students involved in each standardization testing was approximately 65,000. The normative scores--NCE scores, percentile ranks, stanines, grade equivalents, and extended scale scores--can all be derived directly from raw scores.

N-D: The Nelson-Denny Reading Test (N-D), Form E, developed for grades 9 through 12, is also published by Riverside. It is unlike the othernders, and vocational/technical students. Equipercentile equating established TABE norms. The TABE Technical Report reports separate norms for each of the four reference groups.

The norming samples were obtained from institutions across the country that provided the following four types of programs: Adult basic education programs (including literacy and pre-GED), adult offender programs, juvenile offender programs (juveniles 16 years or older) and vocational/technical training programs. A total of 223 institutions from 44 states participated in the norming study. The norming sample consisted of Asians (5.7%), Blacks (44.6%), Hispanics (13.2%), and a variety of others (36.4%). The sample ranged in age from 15-24 (49%) to 65+ years.

TASK: The Stanford Test of Academic Skills (TASK), Level 1, Form E, designed for grades 8 through 12, is published by the Psychological Corporation. It includes eight subtests: Reading Vocabulary, Reading Comprehension, Mathematics, Spelling, English, Social Science, Science, and Using Information. In this study, only the Reading Vocabulary, Reading Comprehension, and Total Reading (the sum of Vocabulary and Comprehension) scores were used. The Reading Vocabulary subtest has an incomplete sentence in the stem, followed by four single-word alternative responses. The Reading Comprehension test uses passages varying from short, quantitative content to relatively long prose.

Norming Groups: The TASK standardization testing was conducted in the Fall, 1981 and Spring, 1982. About 450,000 students in kindergarten through 12th grade took the Stanford Achievement Test (SAT) and the Otis-Lennon School Ability Test. A sample of about 20,000 of the SAT sample was used to vertically equate the SAT and TASK tests to put reading scores on one RGL metric.

The TASK was normed three times during the 1981-82 school year in 300 school districts. This was accomplished through an administration to each grade for which it was designed. Fall testing took place from September 28 to October 16, 1981. Approximately 250,000 pupils participated in the Fall Standardization. The Mid-year standardization was conducted with 15,000 kindergarten and grade 1 pupils from January 25 to February 12, 1982. The Spring standardization took place from April 26 to May 14, 1982 with 200,000 students participating in the program. All students, except kindergartners, took the Otis-Lennon School Ability Test, as well as the 7th edition of the Stanford. The TASK was re-normed in 1986 following a similar procedure.

The samples were chosen to represent the national population in terms of school system enrollment, geographic region, socioeconomic status, and public vs. private school affiliation. A socioeconomic status index (based on median income and percentage of high school degree), served as a stratification variable in selecting the standardization samples. The ethnic background of the sample consisted largely of whites (77%), and also black (13%), or Hispanic (7%).

Table 2 summarizes the information on the reading test norming groups. The six norming groups are widely different.

Table 2

READING TEST NORMS DEVELOPMENT

TEST	DATE (S)	SAMPLE (S)
ABLE	Winter, 1984	515 Adult Prisoners 474 Adult Ed./GEDs 718 Voc/Tech Students
AFRAT	May-July, 1981	5,506 Air Force Recruits 1,049 Army Recruits
G-M	Fall, 1976 Spring, 1977	10th - 12th Grades 86 School Districts
N-D	Fall, 1979	14,000 High School Students 6,000 Two-Year College Students 5,000 Four-Year College Students
TABE	Jan-Mar 1986	1,164 Adult Basic Ed. Students 4,057 Adult Offenders 501 Juvenile Offenders (16+) 613 Voc/Tech Students
TASK	Fall, 1981 Spring, 1982	34,844 9th Graders 24,410 10th Graders 22,414 11th Graders 20,613 12th Graders 1,255 Junior College Freshmen

DEMOGRAPHICS

TEST	% MALE	AGES	Racial			Geographical			
			WH+	BL	HI	NE	MW	SE	W
ABLE	60	38% ≤ 19 38% 20-29 24% 30-50+	71%	24%	5%	20%	38%	28%	15%
AFRAT	85	100% 18-27	50%	50%	NR		Not Reported		
G-M	50	100% Grades 9-12	83%	12%	5%	20%	28%	31%	22%
N-D	50	36% Grades 9-12 27% Two Yr Col 37% Four Yr Col	80%	15%	5%	19%	30%	24%	27%
TABE	71	49% 15-24 31% 25-34 20% 34-65+	42%	45%	13%		Not Reported		
TASK	50	34% 9th 24% 10th 22% 11th 20% 12th 1% Jr Col	81%	13%	6%	19%	26%	27%	28%

Psychometric Characteristics. Table 3 shows the test publishers' reported reliabilities. Subtest and total test reliabilities were generally all high. The reliabilities ranged from .77 to .94.

Table 3

RGL Study Reading Test Reliabilities

Test/Subtest		Rxx/Type	# Items	Test Time	Data Source
ABLE	Vocab	.82 (KR-21)	32	20	Table 16, Page 53, ABLE Booklet.
	Comp	.90 (KR-21)	48	35	
AFRAT	Vocab	.85 (KR-20)	45	15	Personal Communication. Dr. Malcolm Ree, AFHRL, May 1988.
	Comp	.86 (KR-20)	40	35	
	Total	.92 (KR-20)	85	50	
		.94 (ALT)			Study data.
G-M	Vocab	.87 (ALT)	45	20	Table 16, Page 19, G-M Technical Report, 2nd Edition.
		.90 (KR-20)			
	Comp	.81 (ALT)	43	35	
		.90 (KR-20)			
	Total	.90 (ALT)	88	55	
N-D	Vocab	.92 (ALT)	100	15	Page 17, N-D Test Manual.
	Comp	.77 (ALT)	36	20	
	Total	.91 (ALT)	136	35	
TABE	Vocab	.86 (KR-20)	30	17	Table 8, Pages 41-45, TABE Technical Report, 1987.
	Comp	.87 (KR-20)	40	37	
	Total	.91 (KR-20)	70	54	Personal Communication. Dr. Wendy Yen, CTB McGraw-Hill, June 1988.
TASK	Vocab	.87 (ALT)	37	20	Table 6, Page 27, TASK National Norms Booklet.
	Comp	.85 (ALT)	50	30	
	Total	.92 (ALT)	87	50	

RESULTS

Comparison of Reading Groups on Demographics

Table 4 contains information on applicants administered the six reading tests. It supports the effectiveness of the sampling design. More than 3,000 applicants took each reading test, with a range of 3,108 to 3,782 per

instrument. Sex, race, and education level percentages were not significantly different by test. A slight Hispanic overrepresentation in the G-M and TABE and underrepresentation in the ABLE and TASK of about two percent occurred. Ages of examinees were consistent, with a mean age of about 21 and median age of 19. Service representation, as well as reserve/active force membership, were consistent.

The proportion of the applicants taking the ASVAB first was generally a little greater than the optimum 50 percent, probably due to recruiters' preference that applicants take the ASVAB first.

Table 4

Characteristics of Applicants in Reading Test Groups

Variable	ABLE	AFRAT	G-M	N-D	TABE	TASK	TOTAL
Applicants (#)	3,108	3,357	3,558	3,187	3,782	3,430	20,422
Male (%)	80.8	81.5	80.7	81.6	80.9	79.2	80.8
Black (%)	30.6	29.9	24.3	26.2	28.1	29.9	28.1
Hispanic (%)	7.3	4.5	9.8	6.4	9.9	5.9	7.4
Age (Years)							
Mean	20.6	20.6	21.2	20.7	21.2	20.9	20.9
Median	19.0	19.0	19.0	19.0	20.0	19.0	19.0
HS Senior + HSG (%)	86.6	84.8	86.6	86.6	87.4	86.9	86.5
Service (%)							
Army	55.7	53.8	54.6	52.8	55.5	51.8	54.1
Navy	19.9	20.3	20.3	20.9	18.8	20.9	20.2
Marine Corps	8.4	8.3	7.3	9.0	7.2	6.2	7.7
Air Force	14.1	16.0	16.2	15.9	16.5	18.2	16.2
Coast Guard	1.9	1.6	1.5	1.5	2.0	2.9	1.9
Reserve Component(%)	31.1	30.6	34.4	27.9	30.0	30.7	30.8
Reading Test 1st (%)	45.5	46.4	50.5	48.0	47.7	48.2	47.8

Overall, there was no reason to suspect the six reading test groups were not comparable. Nor did the data suggest that the results were not nationally representative of applicants for military service.

Comparison of Reading Groups on AFQT

Table 5 shows the AFQT category distributions of the six reading groups. The percentage represented within each of the five AFQT categories was similar across the six groups. Similarly, the distribution of the total of the groups within the AFQT categories was similar to the combined 1984-86 applicants shown at the bottom of the table. The distribution of 1984-1986 applicants is shown for reference.

Table 5

AFQT Category Distribution of Reading Groups and Applicant Population (Percent)

	AFQT CATEGORY				
Reading Test	I	II	IIIA	IIIB	IV Below
ABLE	4.9	27.3	17.2	26.7	23.9
AFRAT	4.3	25.8	16.5	27.1	26.3
G-M	4.9	27.8	16.8	26.5	24.0
N-D	5.0	27.3	16.5	28.6	22.6
TABE	4.3	27.8	15.9	27.5	24.5
TASK	4.6	25.9	18.1	26.4	24.9
TOTAL	4.6	27.0	16.8	27.1	24.4
1984-86 Active Duty Applicants	4.6	28.5	17.4	26.3	23.1

RGL of Applicants

Table 6 displays the mean RGLs as obtained from the vocabulary and comprehension subtest scores and the total reading scores for each of the six reading tests. Also shown are the RGLs associated with selected percentile points. The six reading tests were relatively similar in average RGL at the 25th percentile, near where many reading ability selection and training remediation decisions would likely be made. AFRAT RGL scores were lower than the other tests at or above the median, probably due to the relatively high ability of Air Force recruits in the AFRAT norming group. The TABE averaged the highest RGL at the median; again likely due to norming group lower ability.

Table 6

Reading Grade Levels of Military Applicants at Various Points in Distributions

Test/Subtest	Mean RGL	Percentiles				
		10	25	50	75	90
Vocabulary						
ABLE	11.1	6.6	8.7	11.1	PHS[2]	PHS[2]
AFRAT	9.6	8.4	8.9	10.3	1.8	2.6
G-M	11.5	7.3	9.0	11.5	PHS[2]	PHS[2]
N-O	12.8	7.1	9.4	12.1	14.3	16.5
TABE	11.9	7.1	9.0	12.9+	12.9+	12.9+
TASK	11.0	6.3	9.1	PHS[2]	PHS[2]	PHS[2]
Comprehension						
ABLE	9.7	6.3	8.2	11.4	PHS[2]	PHS[2]
AFRAT	9.6	7.4	8.7	9.9	11.6	12.4
G-M	10.0	6.1	7.6	10.4	PHS[2]	PHS[2]
N-O	10.6	5.3	7.3	10.7	13.7	16.5
TABE	11.2	8.1	9.2	11.5	12.9+	12.9+
TASK	9.1	5.8	7.4	10.5	PHS[2]	PHS[2]
Total Reading						
ABLE	1	1	1	1	1	1
AFRAT	9.6	8.0	8.7	10.0	11.5	12.5
G-M	10.9	7.0	8.6	11.1	PHS[2]	PHS[2]
N-O	12.0	6.6	8.8	11.3	14.1	16.2
TABE	12.6	7.7	9.1	12.9	12.9+	12.9+
TASK	9.8	6.0	8.0	11.5	PHS[2]	PHS[2]

[1] No ABLE Total Reading score published.
[2] PHS = Post High School.

Selection of ASVAB Composite for Equating

This section consists of: a) descriptive statistics for reading test scores, b) difficulty indices for ASVAB composites and total (where available) reading test scores, and c) multiple regression and factor analyses of ASVAB and reading test scores.

Table 7 displays descriptive statistics for the six reading tests. It shows the N-D Total to have had the opposite skew and the TASK Total with more skew than the other tests. The TASK and the ABLE Comprehension subtests had opposite kurtosis indices from the other tests. The AFRAT and TABE distributions were most alike.

Table 7

Descriptive Statistics for RGL Study Reading Test Raw Scores[1]

Test/ Subtest	N	Mean	S.D.	Skewness	Kurtosis	Median
Vocabulary						
ABLE	3,357	22.21	5.97	- .199	- .837	22.0
AFRAT	3,108	32.41	8.09	- .576	- .288	34.0
G-M	3,558	28.27	8.94	- .043	- .841	31.0
N-D	3,187	43.73	19.59	.683	- .134	40.0
TABE	3,782	22.30	5.98	- .804	- .082	24.0
TASK	3,430	32.53	5.28	- .161	2.408	35.0
Comprehension						
ABLE	3,357	37.38	7.53	- .955	.597	39.0
AFRAT	3,108	29.70	7.55	- .738	- .220	31.0
G-M	3,558	29.59	7.71	- .395	- .552	30.0
N-D	3,187	36.43	13.68	.158	- .601	36.0
TABE	3,782	29.53	6.71	- .642	- .115	30.5
TASK	3,430	38.92	8.32	- .908	.244	41.0
Total Reading						
ABLE [2]	3,357	59.58	12.46	- .521	- .269	61.0
AFRAT	3,108	62.12	14.75	- .612	- .329	64.0
G-M	3,558	57.86	15.68	- .131	- .769	58.0
N-D	3,187	80.16	31.47	.517	- .328	75.0
TABE	3,782	51.83	12.06	- .687	- .173	54.0
TASK	3,430	71.45	12.78	-1.147	.956	75.0

1 TABE uses standard scores.
2 No ABLE Total Reading norms published.

Difficulty Indices

An important comparison in determining acceptability of tests for equating is the similarity of their score distributions. One index of similarity is the mean p-value. Table 8 displays data for the ASVAB composites and reading total scores. P-values for the ASVAB composites are the proportion of examinees who answered the items correctly. Clearly the N-D and G-M were relatively difficult, and the TASK was relatively easy. The TABE, AFRAT, and ABLE had quite similar difficulty indices. From a test difficulty perspective, the latter three reading total scores and the G-M were more appropriate for equating.

Table 8

Mean P-Values for ASVAB Composites and Reading Test Total Scores

ASVAB		READING TEST	
Composite	Mean P-Value	Total Score	Mean P-Value
AFQT-OLD	.711	ABLE	.745
AFQT-NEW	.681	AFRAT	.731
GT	.696	G-M	.658
VE	.735	N-D	.466
VRB	.711	TABE	.740
		TASK	.821

As seen throughout the data presented in this study, the Nelson-Denny Reading Test was not like the other measures of reading. Its difficulty, speededness, correlations with ASVAB subtests, and factor structure made it a poor candidate for equating. Data on the N-D are reported in this section of the paper for completeness, but it was not appropriate as a basis for equating to the ASVAB.

Table 9 shows the intercorrelations between reading scores and ASVAB composites and ASVAB subtests (corrected for unreliability). Differences between coefficients of about .02 are significantly statistically different. In general, reading vocabulary subtests correlated higher with ASVAB composites than comprehension subtests. Correlations were consistently higher for VE and VRB than for the other three composites, approaching unity for both reading subtests and total score. Total reading score correlations were consistently higher than either subsumed subtest. Moreover, as would be expected, the highest correlations of all three reading scores were with WK, PC, and GS subtests. The N-D correlated much higher with ASVAB speeded subtests (NO and CS) than for the other reading tests, confirming the strong speed component in the N-D.

Multiple Regression and Factor Analysis

Table 10 contains the results of six step-wise multiple regressions. The first column indicates the criterion, which in all cases was the reading test total score. The second column indicates the ASVAB subtests which were the predictor variables that are listed in the order in which they were selected into the regression equations. The third column contains the squared multiple correlation as each ASVAB subtest was added. The final column indicates the increased percentage of variance added by each of the ASVAB subtests.

Table 9

CORRECTED INTERCORRELATIONS BETWEEN ASVAB SCORES AND READING TEST SCORES

	AFQT-O	AFQT-M	GT	VE	VRB	GS	AR	MK	PC	NO	CS	AS	MK	MC	EI
ABLE															
VOCAB	0.92	0.95	0.94	1.00*	0.98	0.97	0.78	0.99	0.90	0.34	0.31	0.62	0.75	0.80	0.85
COMP	0.80	0.82	0.82	0.86	0.84	0.76	0.69	0.80	0.89	0.44	0.43	0.45	0.62	0.68	0.65
TOTAL															
AFRAT															
VOCAB	0.90	0.93	0.93	1.00*	0.98	0.93	0.74	1.00*	0.92	0.36	0.37	0.58	0.69	0.76	0.82
COMP	0.85	0.87	0.87	0.91	0.88	0.79	0.74	0.85	0.95	0.45	0.47	0.49	0.65	0.71	0.71
TOTAL	0.89	0.92	0.91	0.98	0.95	0.87	0.74	0.95	0.95	0.41	0.42	0.54	0.72	0.75	0.78
G-M															
VOCAB	0.90	0.92	0.91	1.00*	0.96	0.93	0.74	0.99	0.88	0.37	0.34	0.55	0.68	0.74	0.78
COMP	0.85	0.88	0.86	0.92	0.88	0.82	0.74	0.86	0.90	0.42	0.43	0.43	0.68	0.68	0.67
TOTAL	0.91	0.93	0.92	0.99	0.96	0.89	0.76	0.97	0.92	0.40	0.39	0.51	0.69	0.74	0.76
N-D															
VOCAB	0.88	0.86	0.85	0.92	0.88	0.85	0.69	0.90	0.82	0.44	0.41	0.45	0.68	0.66	0.71
COMP	0.87	0.85	0.83	0.87	0.83	0.76	0.70	0.80	0.90	0.59	0.60	0.31	0.71	0.61	0.60
TOTAL	0.90	0.88	0.86	0.92	0.89	0.84	0.71	0.88	0.88	0.50	0.50	0.40	0.71	0.65	0.69
TABE															
VOCAB	0.88	0.93	0.92	1.00*	0.97	0.90	0.73	0.99	0.93	0.37	0.34	0.50	0.68	0.72	0.74
COMP	0.84	0.88	0.87	0.92	0.88	0.81	0.74	0.87	0.91	0.42	0.41	0.42	0.69	0.70	0.68
TOTAL	0.88	0.92	0.91	0.99	0.95	0.87	0.75	0.95	0.95	0.41	0.39	0.47	0.70	0.72	0.73
TASK															
VOCAB	0.78	0.83	0.83	0.97	0.91	0.82	0.63	0.95	0.87	0.35	0.34	0.47	0.56	0.64	0.67
COMP	0.83	0.85	0.85	0.92	0.88	0.77	0.71	0.84	0.95	0.48	0.48	0.46	0.63	0.71	0.66
TOTAL	0.83	0.88	0.86	0.96	0.91	0.81	0.69	0.91	0.95	0.44	0.44	0.48	0.61	0.71	0.68

* Values calculated at 1.001 - 1.02 due to rounding error.

Table 10

Stepwise Multiple Regression Results

ASVAB Criterion	Correlation Subtest	Percent Squared	Variance[3]
ABLE TOTAL	WK	0.6962	69.6
	PC	0.7496	5.3
	GS	0.7754	2.6
	AR	0.7886	1.3
	MK	0.7899	0.1
	EI	0.7912	0.1
	CS	0.7922	0.1
	MC	0.7926	0.0
AFRAT TOTAL	WK	0.7337	73.4
	PC	0.7891	5.5
	AR	0.8048	1.6
	GS	0.8114	0.7
	CS	0.8130	0.2
	EI	0.8140	0.1
	AS	0.8144	0.0
	MC	0.8148	0.0
G-M TOTAL	WK	0.7418	74.2
	PC	0.7766	3.5
	GS	0.7896	1.3
	AR	0.7962	0.7
	AS	0.7975	0.1
	EI	0.7995	0.2
	MK	0.8000	0.1
	NO	0.8003	0.0
	CS	0.8007	0.0
N-O TOTAL	WK	0.6315	63.2
	MK	0.6776	4.6
	PC	0.7060	2.8
	CS	0.7141	0.8
	GS	0.7217	0.8
	NO	0.7238	0.2
	AS	0.7256	0.2
	EI	0.7280	0.2
	AR	0.7284	0.0
	MC	0.7289	0.0
TABE TOTAL	WK	0.7204	72.0
	PC	0.7702	5.0
	AR	0.7849	1.5
	GS	0.7919	0.7
	AS	0.7956	0.4
	MC	0.7975	0.2
	CS	0.7986	0.1
	MK	0.7990	0.0
TASK TOTAL	WK	0.6702	67.0
	PC	0.7368	6.7
	CS	0.7452	0.8
	GS	0.7514	0.6
	AR	0.7532	0.2
	MK	0.7542	0.1
	MC	0.7553	0.1
	NO	0.7559	0.1

[3]ASVAB subtests contributing significant variance listed.

In each of the six step-wise multiple regression equations, WK entered the regression equation first and accounted for about 70 percent of the variance in reading scores. PC entered next, with the exception of the N-D where MK entered second. WK and PC accounted for 74 to 79 percent of the variance, while 68 percent of the variance was accounted for in the N-D by the WK and MK subtests. Clearly, the overwhelming preponderance of variance in reading total scores was accounted for by WK, followed by PC.

Table 11

Factor Analyses of Reading Total Scores and ASVAB Subtest Scores (Factor loadings >.30)

Test	Eigen-value	Subtest (Loading)					Construct
ABLE	6.22	WK (.81)	ABLE(.47)	GS (.39)	PC (.30)		Verbal
	1.65	CS (.93)	NO (.32)				Speed
	0.72	AS (.91)	EI (.38)	MC (.36)			Technical
	0.60	MK (.85)	AR (.36)				Quantitative
AFRAT	6.25	WK (.86)	AFRAT(.68)	GS (.48)	PC (.39)		Verbal
	1.60	CS (.93)	NO (.32)				Speed
	0.70	AS (.91)	EI (.35)	MC (.33)			Technical
	0.62	MK (.85)	AR (.39)				Quantitative
G-M	6.24	AS (.91)	EI (.36)	MC (.35)			Technical
	1.67	CS (.93)	NO (.33)				Speed
	0.73	WK (.86)	G-M (.78)	GS (.52)	PC (.42)	AR (.31)	Verbal
	0.59	MK (.84)	AR (.38)				Quantitative
N-D	6.10	AS (.92)	EI (.36)	MC (.35)			Technical
	1.67	CS (.93)	NO (.32)				Speed
	0.75	WK (.91)	N-D (.35)	GS (.33)			Verbal
	0.60	MK (.84)	AR (.36)				Quantitative
TABE	6.15	WK (.86)	TABE (.71)	GS (.49)	PC (.41)		Verbal
	1.64	CS (.94)	NO (.32)				Speed
	0.76	AS (.91)	EI (.36)	MC (.34)			Technical
	0.61	MK (.85)	AR (.36)				Quantitative
TASK	6.06	AS (.91)	MC (.36)	EI (.35)			Technical
	1.70	CS (.94)	NO (.33)				Speed
	0.75	WK (.76)	TASK (.37)				Verbal
	0.65	MK (.84)	AR (.37)				Quantitative

The ASVAB subtest scores and reading total scores were factor analyzed. Table 11 displays the results of the factor loadings from six principal components analyses. All six reading tests had a verbal factor on which the total reading test score and the WK and PC subtests were highly loaded. The two exceptions were N-D and TASK in which PC was not represented on the verbal factor. The GS subtest also appeared on the verbal factor in four instances. While it contains a heavy verbal component, it was not considered appropriate for equating to the reading total scores because it does not measure the same abilities as do the WK and PC subtests.

Based upon subtest content, test difficulty, correlation, regression, and factor analytic data shown in Tables 7 through 11, the decision was made that ASVAB-VE was the best anchor test for equating purposes. Thus, VE is used in all further analyses in this paper.

Figure 1 displays the cumulative frequency distributions of the reading tests and ASVAB-VE. Clearly, the AFRAT and TABE best matched the VE distribution.

Equating Reading Tests to ASVAB-VE

Comparison of Standard Errors of Estimate

Table 12 contains the equipercentile equating smoothing methods selected for each reading test.

Table 12

ASVAB VE-Reading Test Equating Smoothing Methods

Reading Test	Equipercentile Equating Method
ABLE VOCAB	LINEAR
AFRAT TOTAL	LINEAR
G-M TOTAL	QUADRATIC
TABE TOTAL[1]	CUBIC
TASK TOTAL	CUBIC

1 Standard score.

Equating Results and Average RGL Across Reading Tests

Table 13 shows the resulting conversion table for converting between VE standard scores and reading test raw scores (or standard scores where applicable) for the five reading tests (without N-D). The final column is the smoothed predicted (linear regression of VE and mean RGL) average RGL across all five reading tests for each of the VE scores where at least two RGL estimates were available. VE standard scores of 25 through 54 provided the regression base. Extrapolations were made from VE=20 through 24 and VE=55 through 62. RGL scores were truncated at less than 2.8 and greater than 12.9, where the reading test publishers had no RGL scale equivalent scores.

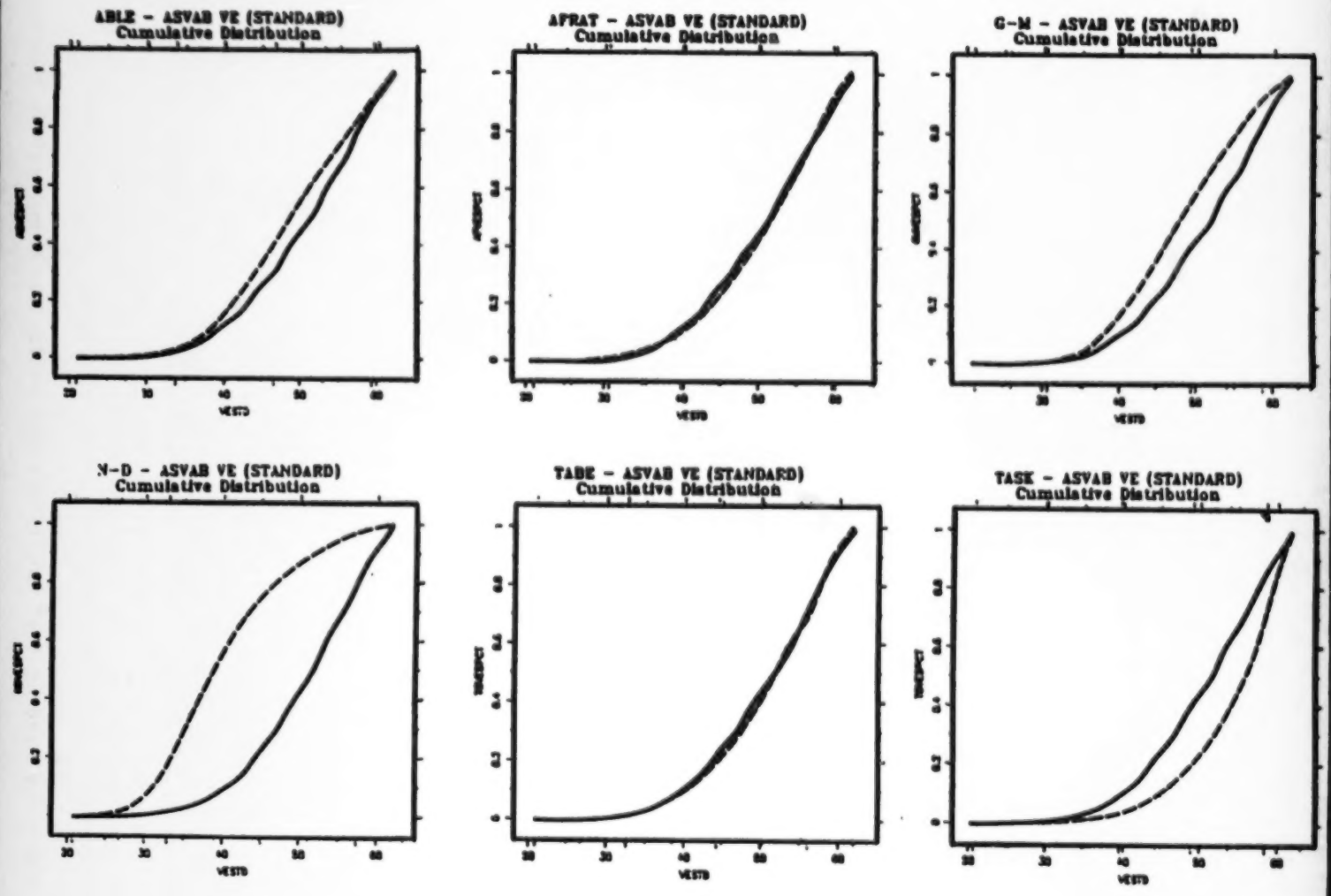

Figure 1. **COMPARISON OF CUMULATIVE FREQUENCY DISTRIBUTION OF SIX READING TEST SCORES AND ASVAB VE STANDARD SCORE**

Table 13

CONVERSION TABLE WITH RECOMMENDED SMOOTHED DOD RGL SCALE

%ILE	ASVAB VE STD	ABLE/VOCAB RAW	ABLE/VOCAB RGL	AFRAT/TOTAL RAW	AFRAT/TOTAL RGL	G-M/TOTAL RAW	G-M/TOTAL RGL	TABE/TOTAL STD	TABE/TOTAL RGL	TASK/TOTAL RAW	TASK/TOTAL RGL	MEAN RGL	SMOOTHED RGL	DoD RGL
1	20												1.3	<2.8
1	21									17	2.7		1.6	<2.8
1	22									18	2.7		1.9	<2.8
1	23									19	2.8		2.2	<2.8
1	24									20	2.9		2.5	<2.8
1	25	4	2.9							21	2.9	2.90	2.8	2.8
1	26	5	3.2	15	4.0					23	3.0	3.40	3.1	3.1
1	27	5	3.2	16	4.2					24	3.1	3.50	3.4	3.4
1	28	6	3.6	18	4.7			666	3.4	26	3.2	3.73	3.7	3.7
1	29	7	3.9	20	6.1			681	4.0	28	3.3	4.33	4.0	4.0
1	30	7	3.9	22	6.7	21	3.8	693	4.6	30	3.5	4.50	4.3	4.3
1	31	8	4.2	24	6.9	22	4.1	701	4.9	32	3.7	4.76	4.6	4.6
1	32	9	4.7	26	7.0	23	4.4	707	5.2	34	3.8	5.02	4.9	4.9
2	33	10	5.1	28	7.1	25	4.8	713	5.4	36	4.0	5.28	5.2	5.2
2	34	10	5.1	30	7.2	26	5.0	718	5.7	38	4.2	5.44	5.5	5.5
3	35	11	5.5	32	7.3	27	5.3	722	5.9	40	4.4	5.68	5.8	5.8
4	36	12	5.8	34	7.5	29	5.6	727	6.2	42	4.6	5.94	6.1	6.1
5	37	12	5.8	36	7.7	30	5.8	731	6.4	45	5.0	6.14	6.4	6.4
7	38	13	6.2	38	7.8	32	6.1	735	6.7	47	5.2	6.40	6.7	6.7
8	39	14	6.6	40	7.9	34	6.5	738	6.9	49	5.5	6.68	7.0	7.0
10	40	15	7.1	42	8.0	35	6.7	742	7.3	51	5.7	6.96	7.3	7.3
12	41	15	7.1	44	8.1	37	7.0	746	7.6	54	6.2	7.20	7.6	7.6
14	42	16	7.6	46	8.2	39	7.3	753	8.2	56	6.6	7.58	7.9	7.9
17	43	17	8.2	48	8.4	41	7.6	756	8.4	58	6.9	7.90	8.2	8.2
20	44	17	8.2	50	8.5	42	7.8	760	8.8	60	7.2	8.10	8.5	8.5
23	45	18	8.7	52	8.7	44	8.2	763	9.0	63	7.8	8.48	8.8	8.8
27	46	19	9.3	54	8.9	46	8.6	767	9.4	65	8.3	8.90	9.1	9.1
30	47	20	9.8	56	9.1	48	9.0	770	10.0	67	8.8	9.34	9.4	9.4
34	48	20	9.8	58	9.3	50	9.5	774	10.8	69	9.3	9.74	9.7	9.7
38	49	21	10.4	59	9.4	52	9.8	777	11.2	71	9.9	10.14	10.0	10.0
40	50	22	11.1	61	9.6	54	10.2	780	12.1	73	10.6	10.80	10.3	10.3
44	51	23	11.8	63	9.8	57	10.9	783	>12.9	74	11.0	10.88	10.6	10.6
49	52	23	11.8	65	10.1	59	11.3	787	>12.9	76	12.0	11.30	10.9	10.9
54	53	24	12.5	67	10.5	61	11.8	791	>12.9	78	>12.9	11.60	11.2	11.2
59	54	25	>12.9	69	10.7	63	12.3	794	>12.9	79	>12.9	11.65	11.5	11.5
63	55	26	>12.9	71	11.0	66	>12.9	798	>12.9	80	>12.9		11.8	11.8
69	56	27	>12.9	73	11.3	68	>12.9	802	>12.9	81	>12.9		12.1	12.1
74	57	28	>12.9	75	11.7	71	>12.9	807	>12.9	82	>12.9		12.4	12.4
80	58	29	>12.9	77	12.2	73	>12.9	811	>12.9	83	>12.9		12.7	12.7
86	59	30	>12.9	79	12.4	76	>12.9	816	>12.9	84	>12.9		13.0	>12.9
91	60	31	>12.9	81	12.6	78	>12.9	822	>12.9	84	>12.9		13.3	>12.9
96	61	32	>12.9	83	12.9	81	>12.9	828	>12.9	85	>12.9		13.6	>12.9
99	62			85	>12.9	84	>12.9	844	>12.9				13.9	>12.9

* REGRESSION SMOOTHED RGL SCALE BASED UPON VE STANDARD SCORES OF 25 THROUGH 54.
DoD RGL SCALE TRUNCATED BELOW 2.8 AND ABOVE 12.9.

Regression Output:

Constant	-4.4715
Std Error of Y Est	0.2700
R Squared	0.9900
No. of Observations	30
Degrees of Freedom	28
x Coefficient(s)	0.3000
Std Error of Coef.	0.0056

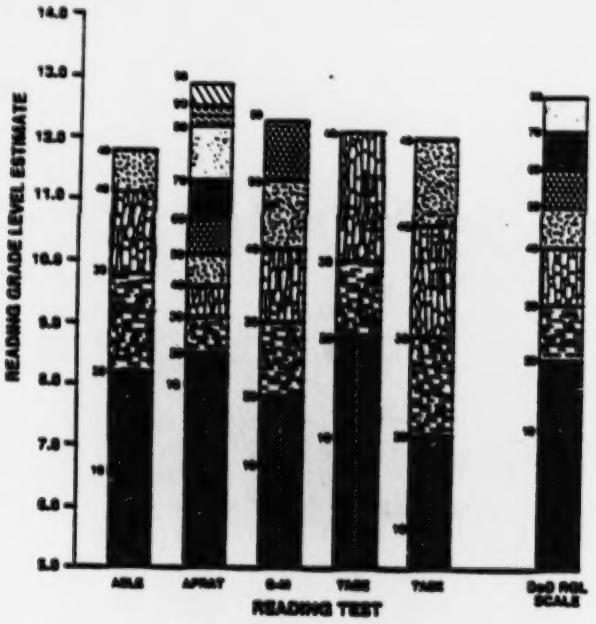

Figure 2. DECILES OF RGLS OF FIVE READING TESTS ALONG WITH PROPOSED DoD RGL SCALE

The RGLs equated to VE scores are reflected in Figures 2 and 3 for each of the five reading tests and the proposed DoD RGL scale. The overall mean RGL for the 20,422 applicants was 10.6. As can be seen, the RGLs associated with the same VE score varied depending on the reading test. At the median, the RGLs were 10.2 for the AFRAT, 11.4 for the G-M, 11.9 for the ABLE, 12.2 for the TASK, and greater than 12.9 for the TABE. A range of three grade levels evolved at the midpoint of the reading test scales! There was no conceptual basis for declaring any of the reading test RGL scales as "right."

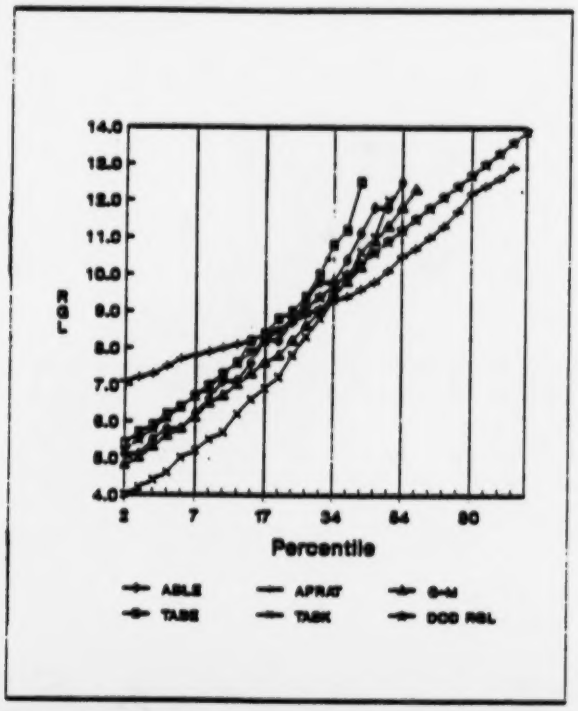

Figure 3. READING GRADE LEVEL CONVERSIONS ASVAB VE STANDARD SCORE AS ANCHOR

Proposed DoD RGL Scale

The fourth objective of this study was to recommend an RGL scale for *use by DoD to report applicant and recruit reading ability to Congress. One* of the primary problems underlying the RGL Study was the lack of consistency between the estimated RGL of a person depending upon which reading test was taken. The results from this study certainly confirmed that condition. RGLs from the publishers' manuals were wildy divergent.

A recommended approach to solving this dilemma was to follow a procedure similar to one used by Mathews and Roach (1983) in which they took the mean RGL of the TABE, G-M, and ABLE at each ASVAB GT scale point to develop the AFRAT RGL scale. Unfortunately, Mathews and Roach were forced to use two of the other reading tests' earlier published data, since samples of AFRAT subjects did not take the ABLE or TABE tests during that data collection.

In the present study, however, this was not a problem. The same population took all six reading tests as well as the ASVAB VE anchor. All reading data were comparable, and only the RGL metric itself was at question. Thus, a recommended DoD RGL scale was developed as follows:

1. Develop conversion table between reading scores and ASVAB VE.

2. Determine RGLs from publishers' test manuals.

3. Compute the mean RGL for each ASVAB VE-equivalent reading test score across reading tests if all reading tests had conversions. This step is shown in the third from the right column, labeled "MEAN RGL", in Table 12.

4. Run a linear regression between ASVAB VE scores and MEAN RGL scores to produce a smoothed RGL metric which reflects an average of all reading test RGL scales. The result from this analysis is shown in the next to the right-hand column of Table 12.

5. Truncate the RGL scale to reflect the publishers' maximum RGL scales and the lower VE Standard points for which no sample data existed.

Proposed DoD Smoothed RGL Scale

The "DoD Smoothed RGL Scale" has several desirable properties. First, it is not directly dependent upon the norming samples of any of the individual reading tests. Thus, particularly high or low ability samples in the reading test data would not drastically distort the resulting RGL scale. Certainly this factor would have been a problem with the relatively high ability AFRAT norming group (primarily Air Force recruits), and perhaps with some or all of the other test samples. Psychometrically, the AFRAT was clearly the best match with the ASVAB for equating purposes; however, its RGL scale would yield much lower estimates of reading ability than the other tests.

Second, the proposed DoD RGL scale is based upon the RGL metrics of five of the major reading tests in the country, yet does not stamp DoD's approval on any of them. The proposed DoD RGL scale would make no value judgement of the "correctness" of any of the scales. It simply establishes a new metric for DoD's use.

Third, all study data were gathered from large samples from a single population. The resulting proposed DoD RGL metric appears to use the best aspects of all data available, both study and test publisher, to produce a single RGL metric.

Fourth, the proposed DoD RGL scale shown in Table 13 has face validity. The median RGL is 10.9; with the scale ranging from grade levels less than 2.8 to greater than 12.9, covering ASVAB VE standard scores from 20 through 62. Breakpoints between the various AFQT Categories were considered in assessing the "credibility" of the scale. The 10th percentile is the minimum AFQT score authorized for entry into the military: this would equate to a DoD RGL of 7th grade-3rd month. The lower boundary of AFQT Category III (average trainability), the 31st percentile, equals an RGL of 9th grade-5th month. Finally, the average 12th grade level reader would score near the 63rd percentile.

CONCLUSIONS AND RECOMMENDATIONS

The RGL Study tested a very large sample of military applicants using six published reading tests and the full ASVAB. It provides a powerful research and analysis tool for the study of reading measurement and the relationship between reading ability testing and aptitude testing. Beyond the specific objectives of this study, the resulting database is a significant contribution to understanding the complex processes of reading and the scores obtained from published reading tests.

Study Conclusions

Results showed that RGL scores emanating from the reading tests yielded widely varying average RGL levels for examinees of equivalent ability. For example, median estimates of military applicant reading ability ranged from the beginning 10th grade level for the AFRAT scores to the end of 12th grade

level for TABE examinees. This result suggests that the RGL scales of different tests yield quite different reading ability estimates for an examinee depending upon which test he or she took. Thus, if DoD reports the reading ability of military applicants or recruits to the Congress, the results will vary by Service since the Services use different measures of reading ability. It was clearly desirable to establish a single metric of reading ability within DoD. This study permitted the conversion of scores from the reading tests to the same scale using an ASVAB composite score as the "anchor" test.

Based upon similar subtest content, test difficulty, reliability, test score distributions, and test score intercorrelations, the ASVAB Verbal (VE) standard score (made up of the Word Knowledge and Paragraph Comprehension subtests) was selected as the best ASVAB anchor test score. Not surprisingly, each of the reading tests also included subtests which purportedly measured knowledge of vocabulary and comprehension of written text.

A thorough evaluation was made to determine the appropriateness of the various reading test scores for equating. The Nelson-Denny Reading Test was judged to be too different from the VE and other reading tests to be used as a basis for equating. It was much more difficult, highly speeded, and less highly correlated with relevant reading scores than the other reading tests. It was, thus, eliminated from consideration as the basis for developing the conversion tables. The Air Force-developed AFRAT and TABE reading tests were most similar to the VE score distribution and psychometric characteristics. The ABLE Vocabulary test, G-M Total Reading, and TASK Total Reading scores were acceptable for equating. Both linear and equipercentile equating methods were evaluated, with equipercentile linear through cubic smoothing methods assessed. The best equating match was between linear equipercentile smoothing of the ASVAB VE and AFRAT Total scores.

A conversion table was developed between each of the five reading test raw scores and the ASVAB VE standard score. The table permits conversion of scores between each of five reading tests and between each of them and the ASVAB VE anchor.

Because none of the five reading tests' RGL scale was demonstrably "better" than the others, selection of one of them as the "DoD RGL scale" did not seem warranted. However, the group of all of them represented a major proportion of the reading tests administered in the United States. Their combined samples provided a powerful basis for generalization to the reading abilities of American youth of enlistment age, certainly stronger than any of the norming groups by themselves. Thus, the study team investigated developing a new DoD RGL scale, based upon the published RGL scales of the five reading tests which were adequate for equating purposes. For each VE score, the equipercentile equivalent raw score from each reading test published RGL was determined. The equipercentile RGLs for the reading tests were then averaged, yielding a mean RGL scale for those VE scores (25 - 54) which included at least two RGL estimates. A linear regression was then run between ASVAB VE and the mean RGL scale to smooth the metric. A full RGL scale was developed by extrapolating the smoothed regression estimates to VE scores from 20 to 24 and 55 to 62. This produced a recommended "DoD RGL

scale" with RGLs ranging from less than 2.8 through greater than 12.9 grade levels. The mean RGL of the military applicant sample (N=20,422) was 10.6. Although somewhat higher than the mean RGL estimated in the Profile of American Youth (DoD, 1983) population (10.1), which used the ABLE as its equating test, the DoD RGL scale mean had "face validity." The 10th percentile, the minimum acceptable aptitude test score for enlistment, equated to an RGL of the 7th grade, 3rd month. The minimum RGL equivalent for AFQT Category III was the 5th month of the 9th grade. Traditionally, a 9th grade reading ability has been the point below which the Services have assigned recruits to remedial reading training.

The proposed DoD RGL scale does not put DoD's stamp of approval on any of the individual RGL scales. It was based upon five samples of over 3,000 each and the entire sample of 20,422 had all taken the ASVAB anchor test in a single study. It would permit each Service to continue to use whichever reading test best met its requirements, with each converting the scores to the DoD RGL scale for Congressional reporting purposes. In the opinion of the study team, the proposed DoD RGL scale had greater credibility than any of the individual reading test RGL metrics.

Study Recommendations

- Make the RGL Study item and test score database available to the Military Service laboratories for continued research and analyses. The size and richness of this research tool is unmatched for the study of reading measurement.
- Implement the RGL conversion table based upon ASVAB VE as the common basis for DoD RGL reporting.
- Implement the proposed "DoD RGL" scale as the metric for future RGL reporting to Congress.

REFERENCES

ABLE Adult Basic Learning Examination norms booklet. (1986). San Antonio, TX: The Psychological Corporation.

Department of Defense. (1988). Population representation in the Military Services: Fiscal year 1987. Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel).

Department of Defense. (1982). Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics).

Eitelberg, M. J., Laurence, J. H., & Waters, B. K., with Perelman, L. S. (1984). Screening for service: Aptitude and education criteria for military entry. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Installations and Logistics).

Gates-MacGinitie Reading Tests teacher's manual. (1978). Chicago, IL: The Riverside Publishing Company.

Grafton, F. C. (1980). Development of a literacy measure for ASVAB 8/9/10. Unpublished manuscript. Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Mathews, J. J., & Roach, B. W. (1983). Reading abilities tests: Development and norming for Air Force use (AFHRL-TR-82-26). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

The Nelson-Denny Reading Test examiner's manual. (1981). Chicago, IL: The Riverside Publishing Company.

TABE Tests of Adult Basic Education forms 5 and 6 technical report. (1987). Monterey, CA: CTB/McGraw-Hill.

TASK Stanford Test of Academic Skills norms booklet. (1983). San Antonio, TX: The Psychological Corporation.

Waters, B. K., Barnes, J. D., Foley, P., Steinhaus, S. D., & Brown, D. C. (1988). Estimating the reading skills of military applicants: Development of an ASVAB to RGL conversion table (HumRRO FR-PRD-88-22). Alexandria, VA: Human Resources Research Organization.